

ABSTRACT

The method of Ermakov and Zolotukhin is discussed along with its later developments. By introducing the idea of pseudo-implementation a practical assessment of the method is made. The performance of the method is found to be unimpressive in comparison with a recent regression method.

CATL 66127 19

On the Method of Ermakov and Zolotukhin for Multiple
Integration

by

R. Cranley and T. N. L. Patterson

I. Introduction

In this paper the method introduced by Ermakov and Zolotukhin [1] and subsequently developed by Ermakov and Granovsky in a series of papers will be discussed. The difficulties of implementation of these methods in all but trivial circumstances have at present prevented practical comparison with other methods. By introducing the idea of pseudo implementation the practical performance of the original method is assessed.

II. Resumé and discussion of theory

The method of Ermakov and Zolotukhin [1] including a generalization given by Ermakov [2] can be briefly described as follows. Consider the multivariate function $g(\underline{x})$ which can be expanded in the region D in terms of the set $\phi_0(\underline{x}), \phi_1(\underline{x}), \dots$, orthonormal with respect to the weight $p(\underline{x})$ according to,

$$g(\underline{x}) = \sum_{i=0}^{\infty} \alpha_i \phi_i(\underline{x}) \quad (1)$$

where,

$$\int_D p(\underline{x}) \phi_i(\underline{x}) \phi_j(\underline{x}) d\underline{x} = \delta_{ij} \quad (2)$$

and

$$\alpha_i = \int_D p(\underline{x}) g(\underline{x}) \phi_i(\underline{x}) d\underline{x}. \quad (3)$$

Ermakov and Zolotukhin [1] and Ermakov [2] define the random variable,

$$\hat{\theta}(\underline{x}_0, \underline{x}_1, \dots, \underline{x}_n) = \frac{w(g, \underline{x}_0, \underline{x}_1, \dots, \underline{x}_n)}{w(\phi_0, \underline{x}_0, \underline{x}_1, \dots, \underline{x}_n)} \quad (4)$$

where,

$$w(g, \underline{x}_0, \underline{x}_1, \dots, \underline{x}_n) = \begin{vmatrix} g(\underline{x}_0) & \phi_1(\underline{x}_0) & \dots & \phi_n(\underline{x}_0) \\ g(\underline{x}_1) & \phi_1(\underline{x}_1) & \dots & \phi_n(\underline{x}_1) \\ \vdots & \vdots & & \vdots \\ g(\underline{x}_n) & \phi_1(\underline{x}_n) & & \phi_n(\underline{x}_n) \end{vmatrix} \quad (5)$$

and show that if $\underline{x}_0, \dots, \underline{x}_n$ are sampled from the density function

$$f(\underline{x}_0, \underline{x}_1, \dots, \underline{x}_n) = w^2(\phi_0, \underline{x}_0, \underline{x}_1, \dots, \underline{x}_n) \prod_{i=0}^n p(\underline{x}_i) / (n+1)! \quad (6)$$

then

$$\theta = \text{mean}(\hat{\theta}) = \int_D p(\underline{x}) \phi_0(\underline{x}) g(\underline{x}) d\underline{x} \quad (7)$$

and

$$\text{var}(\hat{\theta}) = \int_D p(\underline{x}) g^2(\underline{x}) d\underline{x} - \sum_{i=0}^n \alpha_i^2 \quad (8)$$

Only the case $p(\underline{x}) = 1$ has been discussed in the literature. There are indications that $\text{var}(\hat{\theta})$ can be considerably smaller than the variance of the crude Monte Carlo estimator (Handscomb [3]).

The implementation of the method requires sampling to be performed from the multivariate distribution function (6) comprising $(n+1)s$ variables, where s is the number of dimensions of the integral.

Such sampling could in principle be performed by the rejection technique (Handscorn [3]) or by the use of conditional probabilities (Schreider [9]) as suggested by the original authors. There is little doubt that the former would be enormously inefficient while the latter would require the analytical evaluation of a very large number of multiple integrals involving the orthonormal basis functions which would generally be intractable. Admittedly the sampling distribution (6) does not depend on the integrand so that it might be possible to proceed by having tables of random points available as suggested by Ermakov [2]. However, the distribution does depend on the orthonormal basis, on the value of the parameter n and on the dimension of the integrand so that the tables would indeed be extensive. Additionally the dependence of the sampling distribution on n would necessitate the discarding of all previous computational labour should n be changed.

The problem has been discussed further in a series of papers by Ermakov and others. It is shown that in principle the sampling distribution of the general method can be modified to give a reduction in variance for square summable functions (Ermakov [4]) when the number of orthonormal functions, n , equals the multiplicity of the distribution function N . The complexity of the sampling problem increases.

The case $n < N$ is discussed by Ermakov [2]. It is shown that a variance reduction can be achieved although again the sampling problem increases in complexity and a least squares analysis has to be performed. The case $n > N$ is also discussed by Ermakov [5]. The variance can be reduced in this case by imposing relations between the nodes of

the density function. For example, if in the 2-dimension case the orthonormal system of functions is taken to be a product of two systems of one dimensional orthonormal systems and the sample points are confined to a 2-dimensional grid then the variance of the resulting interpolatory estimator is lower than that which would be obtained were the points unrestricted. In this case the sampling problem reduces the selecting points from two multivariate scalar functions, still a very considerable problem. However this reintroduces the difficulties associated with the rapid growth of the number of integrand evaluations.

Following the approach of Ermakov [5], Granovsky [6] discussed in detail random quadratures of the so-called Gaussian type which impose a maximum number of relationships between the nodes of the Ermakov and Zolotukhin general method. The sampling distribution depends then only on a single node with consequent simplification. It is interesting to note from one of Granovsky's examples,

$$\int_{-1}^1 g(x) dx, \quad (9)$$

with the orthonormal functions,

$$\phi_0(x) = \frac{1}{\sqrt{2}} \quad \phi_1(x) = \frac{\sqrt{3}}{\sqrt{2}} x \quad (10)$$

the Gaussian type estimator is,

$$\hat{\theta} = f(x) + f(-x) \quad (11)$$

with x uniformly sampled, viz. the simplest form of the antithetic variate method. Ermakov and Zolotukhin in this case would obtain

$$\hat{\theta} = \sqrt{2} \frac{x_1 f(x_0) - x_0 f(x_1)}{(x_1 - x_0)} \quad (12)$$

where x_0 and x_1 are jointly sampled from

$$w_2^2(x_0, x_1) = \frac{3}{8} (x_1 - x_0)^2 \quad (13)$$

By introducing the generalized Haar functions as the orthonormal basis Ermakov ([4], [7]) gives an elegant description of stratified sampling. This discussion has also been given independently by Handscomb [3]. As is well known the variance of these stratified estimators is always an improvement on the Crude Monte Carlo.

In connection with sampling from the improved densities, referred to as admissible randomized experimental designs, Granovsky and Ermakov [8], it is noted that the problem can be reduced to a sequence of linear programming problems provided that the density function is defined on a finite set of points. In practice this can always be made the case. However, admissible randomized experimental designs have only been shown to exist under special circumstances, for example, when the orthonormal system is the Haar system. Ermakov [6] discussed this as a trivial case of his general discussion.

In the light of the difficulties associated with the implementation of these general methods, it would be of interest to gain some feeling of general performance in practice. In contrast to the original work of Ermakov and Zolotukhin [1], which provides closed forms for the

quadrature rules, the later work (for example, Granovsky [6]) while theoretically illuminating, fails at present to provide such forms. In further discussion attention will thus be confined to assessment of the original method.

III. General Discussion

In this section a number of points will be noted which must be taken into account when assessing any Monte-Carlo method.

A common way of comparing two methods is by the accuracy achieved with a given number of integrand evaluations. In general if the integrand is of great computational complexity this means of comparison will be useful. However if the integrand is of fairly simple form the complexity of the logic of the method may dominate the computational labour so that integrand evaluations would not fairly compare methods. In this paper it will be assumed that the former situation holds, that is computational labour will be measured by the number of integrand evaluations.

So called improved Monte Carlo methods are characterized by their attempts to reduce the variance of Crude Monte Carlo estimator. This can be achieved either implicitly or explicitly. Importance sampling and the control variate methods typify the explicit reduction and an automatic scheme using Bernstein polynomials has recently been introduced

by Rosenberg [10]. The antithetic variate method of Hammersley and Mauldon [11] and Cranley and Patterson [1]] typify implicit reduction.

Explicit methods proceed by directly transforming the integrand to produce a new estimator which hopefully has smaller variance. The labour invested in doing this is not utilized in subsequent sampling of the new estimator. The variance reduction can in fact be minimal and examples which show this can easily be constructed. This can be especially serious in methods such as Rosenberg's where this labour involves evaluations of the original integrand. In contrast, in the implicit methods the process of variance reduction yields an estimate of the integral without further sampling.

Since the variance reduction normally costs some integrand evaluations it is important in comparing methods to consider not the variance of the new estimator but rather the variance of its mean with the same investment in integrand evaluations. For example the Ermakov and Zolotukhin estimator has variance given by (8) but uses $n+1$ integrand evaluations per sample point. If the mean of $\hat{\theta}$ were evaluated with m sets of random vectors then $(n+1)m$ integrand evaluations would be required. For the case $\phi_0(\underline{x}) = 1$ and $p(\underline{x}) = 1$ it is easy to show that with this investment in integrand evaluations the variance of the mean of the Crude Monte Carlo estimator will be smaller than that of the Ermakov and Zolotukhin estimator when the integrand $g(x)$ satisfies,

$$\sum_{i=1}^n \alpha_i^2 < \frac{n}{n+1} \text{ var } (g) \quad (14)$$

It is not difficult to achieve this condition. The simple function

$$g(x) = a \cos 2p \cos^{-1} x + \sin q \cos^{-1} x \quad (15)$$

for x in $[-1, 1]$ satisfies this requirement where $q > 2p + 1$,

$n \leq 2p$ in (4) and

$$a^2 < \frac{pq^2}{4(4q^2-1)} - \frac{(16p^2-1)(4p^2-1)^2}{p^4(8p^2-5)} \quad (16)$$

IV. Pseudo-implementation

The difficulties of implementation of the Ermakov and Zolotukhin method have been discussed by Handscomb [3] and only in fairly trivial circumstances can the performance of the method be directly assessed. However, in performing test integrations one usually has information which could not in other circumstances be regarded as being available - for example, the exact value of the integral! Taking this a step further it is natural to ask if a knowledge of the Fourier expansion coefficients of the test integrand could be used to predict the results that would be obtained were the method directly implemented. It is not difficult to see how this can be done.

Consider the estimator,

$$\tilde{\theta}(\underline{x}) = \{g(\underline{x}) - \sum_{i=1}^n \alpha_i \phi_i(\underline{x})\} / \phi_0(\underline{x}) \quad (17)$$

where α_i are the Fourier coefficients of $g(\underline{x})$ and where \underline{x} is sampled from the density function

$$f(\underline{x}) = p(\underline{x}) \phi_0^2(\underline{x}) \quad (18)$$

It is easily shown that,

$$\text{mean } (\tilde{\theta}) = \int_D p(\underline{x}) \phi_0(\underline{x}) g(\underline{x}) d\underline{x} \quad (19)$$

and

$$\text{var } (\tilde{\theta}) = \int_D p(\underline{x}) g^2 d\underline{x} - \sum_{i=0}^n \alpha_i^2 \quad (20)$$

which are the same as for the Ermakov and Zolotukhin estimator (4).

Furthermore, $\tilde{\theta}$, by definition, integrates exactly any linear combination of ϕ_0, \dots, ϕ_n . Thus $\tilde{\theta}$ embodies the properties of (4) and hence can be used to indirectly assess the Ermakov and Zolotukhin method. This procedure will be termed pseudo-implementation. If attention is restricted to the case when $p(\underline{x}) = 1$ and $\phi_0(\underline{x}) = 1$ then the pseudo-implementation becomes particularly simple since \underline{x} is then sampled from the uniform distribution.

V. Experiments

In this section some results of pseudo-implementation of the Ermakov and Zolotukhin method will be described. As a basis for judgment of these results comparison will be made with a regression method for Monte Carlo integration described by Cranley and Patterson [12] which falls into the category of an implicit variance reduction method. At this point a brief review of this method will be given.

Attention is restricted to the evaluation of integrals of the form,

$$\theta = 2^{-s} \int_{-1}^1 dx_1 \dots \int_{-1}^1 dx_s g(x_1, \dots, x_s). \quad (21)$$

Consider the estimator of θ defined by,

$$\hat{\theta} = \sum_{i=1}^m a_i T_i(\underline{x}) \quad (22)$$

which is a weighted sum of basic unbiased estimators of θ , $T_i(\underline{x})$. The estimator will be unbiased when,

$$\sum_{i=1}^m a_i = 1 \quad (23)$$

The a_i and θ^* , the estimated mean of $\hat{\theta}$, are then calculated such that,

$$\sum_{j=1}^N \left\{ \sum_{i=1}^m a_i T_i(\underline{x}_j) - \theta^* \right\}^2 = \text{minimum} \quad (24)$$

a straightforward linear least squares problem. The \underline{x}_j are uniformly sampled. For simplicity attention is confined to the one dimensional case. Eq. (22) will be referred to as estimator C when the $T_i(x)$ are defined by

$$T_i(x) = \frac{1}{i} \sum_{j=1}^i g\left(\frac{x - i + 2j - 1}{i}\right) \quad (25)$$

and as estimator D when $T_i(x)$ are defined by,

$$T_{2i}(x) = \frac{1}{2i} \sum_{j=1}^{2i} g\left(\frac{x - 2i + 2j - 1}{2i}\right) \quad (26)$$

and

$$T_{2i+1}(x) = \frac{1}{2i+1} \sum_{j=1}^{2i+1} g\left(\frac{-x - 2i + 2j - 2}{2i+1}\right) \quad (27)$$

$$i = 1, 2, \dots, \left[\frac{m}{2}\right].$$

It can be shown that C will integrate exactly all polynomials of degree $m-1$ using $m(m+1)/2$ integrand evaluations per sample point while D integrates polynomials of degree $2m-1$ using $m(m+1)$ integrand evaluations per sample point. When a polynomial basis is chosen, the Ermakov and Zolotukhin method will integrate exactly the same class of functions. Several other choices of T_i are discussed by Cranley and Patterson [12].

Estimator D has some useful and desirable properties. It reduces the variance implicitly and so can be applied automatically. It requires sampling only from the uniform distribution. No integrand evaluations need be lost if either the sample size N or the number of basic estimators m is increased.

For computational simplicity attention is limited to some one dimensional integrals which have been considered in the literature and to a polynomial basis. Apart from calculating the multivariate orthonormal system, the pseudo-implementation could as easily be applied to a multidimensional case. However, since the methods being compared can integrate the same class of functions it would not be expected that there would be any loss of generality.

The test integrals are given in Table 1. The first four were chosen from Davis and Rabinowitz [13] and have since been used extensively in tests. The fifth example has infinite variance which would lead to a breakdown in the theory for both methods. Unlike the first four integrands whose Fourier coefficients were computed numerically, it has a convenient Fourier expansion as does example six.

In all cases and for each method the variance was estimated using a very conservative sample size of 80. For the first four examples the basic regression estimator D was used with 20 function evaluations per sample point and has degree 7. The Ermakov and Zolotukhin estimator with the same number of evaluations was used and has degree 19°. For the last two examples the regression estimator C was used with 21 integrand evaluations per sample point and has degree 10. The Ermakov and Zolotukhin estimator with degree 39 was used in this case. Twice as many orthonormal functions were used to compensate for the fact that the functions were even.

It is clear from the Table that the Ermakov and Zolotukhin method does not emerge with any advantage over the easily implementable regression scheme. Indeed, only in one case does a significant improvement appear. Generally the regression method is better by several orders of magnitude.

It would appear from these rather limited experiments that the improvements to the basic method of Ermakov and Zolotukhin would have to be considerable to compete with the regression method. Although it could be argued that a more suitable basis than the polynomial one could have been chosen, in practice this would mean having extensive tables of random points available from the relevant distributions. In any case there is no obvious choice of basis for the general integrand.

TABLE 1

Integration of $\int_0^1 f(x)dx$ by regression (R) and the Ermakov
Zolotukhin (EZ) method

Function	Error		Std. dev.	
	R	EZ	R	EZ
$x^{1/2}$	$1.2(-9)^*$	$4.2(-6)$	$1.3(-9)$	$2.2(-5)$
$x^{3/2}$	$7.8(-11)$	$3.6(-8)$	$1.6(-10)$	$1.1(-7)$
$(1+0.5 \sin 10\pi x)^{-1}$	$1.1(-6)$	$5.9(-3)$	$2.1(-6)$	$1.2(-2)$
$(1+x^4)^{-1}$	$3.8(-10)$	$2.8(-13)$	$1.3(-9)$	$3.4(-12)$
$[1-(2x-1)^2]^{1/2}$	$3.9(-13)$	$1.5(-5)$	$3.5(-13)$	$1.9(-5)$
$[1-(2x-1)^2]^{-1/2}$	$9.6(-10)$	$2.8(-2)$	$1.1(-10)$	$3.3(-2)$

* $1.2(-9)$ denotes 1.2×10^{-9} .

REFERENCES

1. S. M. Ermakov and V. G. Zolotukhin, Polynomial approximations and the Monte Carlo method, *Theor. Prob. Appl.*, 5 (1960), pp 428-431.
2. S. M. Ermakov, On estimation of parameters for indirect measurements, *Proc. Steklov Inst. Maths.*, 79 (1965), pp 117-124.
3. D. C. Handscomb, Remarks on a Monte Carlo integration method, *Num. Maths.*, 6 (1964), pp 261-268.
4. S. M. Ermakov, Admissibility procedures of the Monte Carlo method, *Soviet Math. Dokl.*, 8 (1967), pp 66-67.
5. S. M. Ermakov, Random quadratures of improved accuracy, U.S.S.R. *Comput. Math. & Math. Phys.*, 3 (1964), pp 213-219.
6. B. L. Granovsky, On random quadratures of Gauss type, *Zh. Vychisl. Mat. i Mat. Fiz.*, 8 (1968), pp 879-884.
7. S. M. Ermakov, Interpolation with respect to random points, U.S.S.R. *Comp. Math. & Math. Phys.*, 3 (1963), pp 243-249.
8. B. L. Granovsky & S. M. Ermakov, A nonparametric approach to the problem of designing regression experiments, *Soviet Math. Dokl.*, 9 (1968), pp 611-613.
9. Yu. A. Schreider, The Monte Carlo Method, Pergamon Press, 1966.
10. L. Rosenberg, Bernstein polynomials and Monte Carlo integration, *SIAM J. Numer. Anal.*, 4 (1967), pp 566-574.

11. J. M. Hammersley & J. G. Mauldon, General principles of antithetic variates, Proc. Camb. Phil. Soc., 52 (1956), pp 476-481.
12. R. Cranley & T. N. L. Patterson, A regression method for the Monte Carlo evaluation of multidimensional integrals, Numer. Math., 16 (1970), pp 58-72.
13. P. J. Davis & P. Rabinowitz, Numerical integration, (1967), Blaisdell Publishing Co.